GenCore version 4.5

OM protein - protein search, using sw model

Run on: December 27, 2001, 16:53:19 ; Search time 41.26 Seconds
(without alignments)
420.096 Million cell updates/sec

Title: US-09-830-647-2
Perfect score: 1206
Sequence: 1 MNSGAMRIHSKGHFQGGIQV.....LKKPFVKVEDMSQSPAVHLM 234

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1206	100.0	234	21	AAB03759	Human H37 amino ac
2	1170	97.0	674	21	AAB42482	Human ORFX ORF2246
3	1170	97.0	674	21	AAB03758	Human H37 amino ac
4	308	25.5	60	22	AAM19849	Peptide #6283 enco
5	308	25.5	60	22	AAM33333	Peptide #7370 enco
6	281	23.3	55	22	AAM32811	Peptide #6848 enco
7	259	21.5	49	22	AAM18938	Peptide #5372 enco
8	259	21.5	49	22	AAM31512	Peptide #5549 enco
9	208	17.2	170	22	AAB95297	Human protein sequ
10	96	8.0	257	22	AAU03696	Group B Streptococ
11	93.5	7.8	1435	20	AAY49070	PolC gene product
12	93.5	7.8	1435	22	AAB31934	Amino acid sequenc
13	89	7.4	1087	20	AAY19935	B. burgdorferi ant
14	89	7.4	1119	20	AAY19934	B. burgdorferi ant
15	88	7.3	700	12	AAR13227	Novel endoglucanas
16	87.5	7.3	1442	22	AAG82479	S. epidermidis ope
17	87	7.2	617	12	AAR15241	B.lautus endol cor
18	86	7.1	947	19	AAW81168	Transcriptional re
19	86	7.1	947	20	AAY07114	WO9904265 Seq ID N
20	85	7.0	536	21	AAY76007	Murine RIP protein
21	85	7.0	536	22	AAB55946	Skin cell protein,
22	85	7.0	714	21	AAY70209	Murine TANGO 130 p
23	85	7.0	976	17	AAW02289	Mouse neuron restr
24	84.5	7.0	299	16	AAR75416	Rat regucalcin, a
25	84.5	7.0	343	21	AAG60518	Arabidopsis thalia
26	84.5	7.0	347	22	AAG81684	S. epidermidis ope
27	84.5	7.0	347	22	AAG82087	S. epidermidis ope
28	84.5	7.0	412	20	AAY49151	Amino acid sequenc
29	84.5	7.0	425	20	AAY49248	N-terminal region
30	84.5	7.0	425	20	AAY32187	N-terminal choline
31	84.5	7.0	887	22	AAM39431	Human polypeptide
32	84.5	7.0	2781	21	AAY57453	Human transcriptio
33	84.5	7.0	2907	21	AAY57452	Human transcriptio
34	84	7.0	590	21	AAY76123	Murine RIP protein
35	84	7.0	590	22	AAB56062	Skin cell protein,
36	84	7.0	763	21	AAY79154	Mouse protein kina
37	84	7.0	786	21	AAY69163	Amino acid sequenc
38	84	7.0	787	21	AAY76079	Murine protein kin
39	84	7.0	787	22	AAB56018	Skin cell protein,
40	83.5	6.9	317	21	AAG21891	Arabidopsis thalia
41	83	6.9	996	22	AAM39322	Human polypeptide
42	83	6.9	1023	22	AAM41108	Human polypeptide
43	82.5	6.8	273	18	AAW20667	H. pylori cytoplas
44	82.5	6.8	300	18	AAW20098	H. pylori cytoplas
45	82.5	6.8	746	22	AAG81779	S. epidermidis ope

ALIGNMENTS

RESULT 1
AAB03759
ID AAB03759 standard; Protein; 234 AA.
XX
AC AAB03759;
XX
DT 04-OCT-2000 (first entry)
XX
DE Human H37 amino acid sequence #2.
XX
KW H37; human; Cdc7 regulatory subunit; cytostatic; proliferative; cancer;
KW anti-proliferative; replication regulator; stem cell.
XX
OS Homo sapiens.
XX
PN WO200026250-A1.
XX
PD 11-MAY-2000.
XX
PF 01-NOV-1999; 99WO-JP06076.
XX
PR 30-OCT-1998; 98JP-0311408.
XX
PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.
PA (ARAI/) ARAI K.
PA (MASA/) MASAI H.
XX
PI Arai K, Masai H;
XX
DR WPI; 2000-365580/31.
DR N-PSDB; AAA53484.
XX
PT Human H37 proteins with a Cdc7 activity regulatory subunit, for
PT controlling cell replication and cell proliferation, useful in treating

PT cancers and diseases due to abnormal production of stem cells -
XX
PS Claim 2; Page 46-47; 55pp; Japanese.
XX
CC The present sequence represents a human H37 protein sequence. H37 is a
CC protein with a Cdc7 activity regulatory subunit. The invention relates to
CC two H37 protein and nucleotide sequences. H37 proteins exhibit
CC cytostatic, proliferative, anti-proliferative, and cell replication
CC regulatory activities. The proteins, encoded genes and DNA fragments are
CC useful in treating cancers and other diseases resulting from abnormal
CC production of stem cells. Antibodies directed against one of the H37
CC proteins can be used to inhibit cell proliferation.
XX
SQ Sequence 234 AA;

Query Match 100.0%; Score 1206; DB 21; Length 234;
Best Local Similarity 100.0%; Pred. No. 1.6e-113;
Matches 234; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MNSGAMRIHSKGHFQGGIQVKNEKNRPSLKSLKTDNRPEKSKCKPLWGKVFYLDLPSVTI 60
Db 1 mnsgamrihskghfqggiqvkneknrpslkslktdnrpekskckplwgkvfyldlpsvti 60

Qy 61 SEKLQKDIKDLGGRVEEFLSKDISYLISNKKEAKFAQTLGRISPVPSPESAYTAETTSPH 120
Db 61 seklqkdikdlggrveeflskdisylisnkkeakfaqtlgrispvpspesaytaettsph 120

Qy 121 PSHDGSSFKSPDTVCLSRGKLLVEKAIKDHDFIPSNSILSNALSWGVKILHIDDIRYYIE 180
Db 121 pshdgssfkspdtvclsrgkllvekaikdhdfipsnsilsnalswgvkilhiddiryyie 180

Qy 181 QKKKELYLLKKSSTSVRDGGKRVGSGAQKTRTGRLKKPFVKVEDMSQSPAVHLM 234
Db 181 qkkkelyllkksstsvrdggkrvgsgaqktrtgrlkkpfvkvedmsqspavhlm 234

RESULT 2
AAB42482
ID AAB42482 standard; Protein; 674 AA.
XX
AC AAB42482;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF2246 polypeptide sequence SEQ ID NO:4492.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach M;
XX
DR WPI; 2000-602362/57.
DR N-PSDB; AAC76691.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 11; Page 3676-3677; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 674 AA;

Query Match 97.0%; Score 1170; DB 21; Length 674;
Best Local Similarity 100.0%; Pred. No. 3.1e-109;
Matches 227; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MNSGAMRIHSKGHFQGGIQVKNEKNRPSLKSLKTDNRPEKSKCKPLWGKVFYLDLPSVTI 60
Db 1 mnsgamrihskghfqggiqvkneknrpslkslktdnrpekskckplwgkvfyldlpsvti 60

Qy 61 SEKLQKDIKDLGGRVEEFLSKDISYLISNKKEAKFAQTLGRISPVPSPESAYTAETTSPH 120
Db 61 seklqkdikdlggrveeflskdisylisnkkeakfaqtlgrispvpspesaytaettsph 120

Qy 121 PSHDGSSFKSPDTVCLSRGKLLVEKAIKDHDFIPSNSILSNALSWGVKILHIDDIRYYIE 180
Db 121 pshdgssfkspdtvclsrgkllvekaikdhdfipsnsilsnalswgvkilhiddiryyie 180

Qy 181 QKKKELYLLKKSSTSVRDGGKRVGSGAQKTRTGRLKKPFVKVEDMSQ 227
Db 181 qkkkelyllkksstsvrdggkrvgsgaqktrtgrlkkpfvkvedmsq 227

RESULT 3
AAB03758
ID AAB03758 standard; Protein; 674 AA.
XX
AC AAB03758;
XX
DT 04-OCT-2000 (first entry)
XX
DE Human H37 amino acid sequence #1.
XX
KW H37; human; Cdc7 regulatory subunit; cytostatic; proliferative; cancer;
KW anti-proliferative; replication regulator; stem cell.

XX
OS Homo sapiens.
XX
PN WO200026250-A1.
XX
PD 11-MAY-2000.
XX
PF 01-NOV-1999; 99WO-JP06076.
XX
PR 30-OCT-1998; 98JP-0311408.
XX
PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.
PA (ARAI/) ARAI K.
PA (MASA/) MASAI H.
XX
PI Arai K, Masai H;
XX
DR WPI; 2000-365580/31.
DR N-PSDB; AAA53483.
XX
PT Human H37 proteins with a Cdc7 activity regulatory subunit, for
PT controlling cell replication and cell proliferation, useful in treating
PT cancers and diseases due to abnormal production of stem cells -
XX
PS Claim 1; Fig 5; 55pp; Japanese.
XX
CC The present sequence represents a human H37 protein sequence. H37 is a
CC protein with a Cdc7 activity regulatory subunit. The invention relates to
CC two H37 protein and nucleotide sequences. H37 proteins exhibit
CC cytostatic, proliferative, anti-proliferative, and cell replication
CC regulatory activities. The proteins, encoded genes and DNA fragments are
CC useful in treating cancers and other diseases resulting from abnormal
CC production of stem cells. Antibodies directed against one of the H37
CC proteins can be used to inhibit cell proliferation.
XX
SQ Sequence 674 AA;

Query Match 97.0%; Score 1170; DB 21; Length 674;
Best Local Similarity 100.0%; Pred. No. 3.1e-109;
Matches 227; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 MNSGAMRIHSKGHFQGGIQVKNEKNRPSLKSLKTDNRPEKSKCKPLWGKVFYLDLPSVTI 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 mnsgamrihskghfqggiqvkneknrpslkslktdnrpekskckplwgkvfyldlpsvti 60

Qy     61 SEKLQKDIKDLGGRVEEFLSKDISYLISNKKEAKFAQTLGRISPVPSPESAYTAETTSPH 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 seklqkdikdlggrveeflskdisylisnkkeakfaqtlgrispvpspesaytaettsph 120

Qy    121 PSHDGSSFKSPDTVCLSRGKLLVEKAIKDHDFIPSNSILSNALSWGVKILHIDDIRYYIE 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 pshdgssfkspdtvclsrgkllvekaikdhdfipsnsilsnalswgvkilhiddiryyie 180

Qy    181 QKKKELYLLKKSSTSVRDGGKRVGSGAQKTRTGRLKKPFVKVEDMSQ 227
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    181 qkkkelyllkksstsvrdggkrvgsgaqktrtgrlkkpfvkvedmsq 227
```

RESULT 4
AAM19849
ID AAM19849 standard; Protein; 60 AA.
XX
AC AAM19849;
XX
DT 12-OCT-2001 (first entry)
XX
DE Peptide #6283 encoded by probe for measuring cervical gene expression.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 27; SEQ ID No 24675; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP: see AAI10068-AAI28459). The present sequence is a peptide encoded
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 60 AA;

Query Match 25.5%; Score 308; DB 22; Length 60;
Best Local Similarity 100.0%; Pred. No. 8.8e-24;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy     74 RVEEFLSKDISYLISNKKEAKFAQTLGRISPVPSPESAYTAETTSPHPSHDGSSFKSPDT 133
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 rveeflskdisylisnkkeakfaqtlgrispvpspesaytaettsphpshdgssfkspdt 60
```

RESULT 5
AAM33333
ID AAM33333 standard; Protein; 60 AA.
XX
AC AAM33333;
XX
DT 17-OCT-2001 (first entry)
XX
DE Peptide #7370 encoded by probe for measuring placental gene expression.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.

```
PR  30-JUN-2000; 2000US-0608408.
PR  03-AUG-2000; 2000US-0632366.
PR  21-SEP-2000; 2000US-0234687.
PR  27-SEP-2000; 2000US-0236359.
PR  04-OCT-2000; 2000GB-0024263.
XX
PA  (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI  Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR  WPI; 2001-488897/53.
XX
PT  Human genome-derived single exon nucleic acid probes useful for
PT  analyzing gene expression in human placenta -
XX
PS  Claim 27; SEQ ID No 33602; 654pp; English.
XX
CC  The present invention relates to single exon nucleic acid probes (SENP:
CC  see AAI31315-AAI57546). The present sequence is a peptide encoded by one
CC  such probe. The probes are useful for producing a microarray for
CC  predicting, measuring and displaying gene expression in samples derived
CC  from human placenta. The probes are useful for antenatal diagnosis of
CC  human genetic disorders.
XX
SQ  Sequence   60 AA;


  Query Match             25.5%;  Score 308;  DB 22;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 8.8e-24;
  Matches   60;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 RVEEFLSKDISYLISNKKEAKFAQTLGRISPVPSPESAYTAETTSPHPSHDGSSFKSPDT 133
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 rveeflskdisylisnkkeakfaqtlgrispvpspesaytaettsphpshdgssfkspdt 60


RESULT   6
AAM32811
ID  AAM32811 standard; Protein; 55 AA.
XX
AC  AAM32811;
XX
DT  17-OCT-2001 (first entry)
XX
DE  Peptide #6848 encoded by probe for measuring placental gene expression.
XX
KW  Probe; microarray; human; placenta; antenatal diagnosis;
KW  genetic disorder.
XX
OS  Homo sapiens.
XX
PN  WO200157272-A2.
XX
PD  09-AUG-2001.
XX
PF  30-JAN-2001; 2001WO-US00663.
XX
PR  04-FEB-2000; 2000US-0180312.
PR  26-MAY-2000; 2000US-0207456.
PR  30-JUN-2000; 2000US-0608408.
PR  03-AUG-2000; 2000US-0632366.
PR  21-SEP-2000; 2000US-0234687.
PR  27-SEP-2000; 2000US-0236359.
PR  04-OCT-2000; 2000GB-0024263.
XX
PA  (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI  Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR  WPI; 2001-488897/53.
XX
PT  Human genome-derived single exon nucleic acid probes useful for
PT  analyzing gene expression in human placenta -
XX
PS  Claim 27; SEQ ID No 33080; 654pp; English.
XX
CC  The present invention relates to single exon nucleic acid probes (SENP:
CC  see AAI31315-AAI57546). The present sequence is a peptide encoded by one
CC  such probe. The probes are useful for producing a microarray for
CC  predicting, measuring and displaying gene expression in samples derived
CC  from human placenta. The probes are useful for antenatal diagnosis of
CC  human genetic disorders.
XX
SQ  Sequence   55 AA;


  Query Match             23.3%;  Score 281;  DB 22;  Length 55;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-21;
  Matches   55;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 RVEEFLSKDISYLISNKKEAKFAQTLGRISPVPSPESAYTAETTSPHPSHDGSSF 128
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 rveeflskdisylisnkkeakfaqtlgrispvpspesaytaettsphpshdgssf 55


RESULT   7
AAM18938
ID  AAM18938 standard; Protein; 49 AA.
XX
AC  AAM18938;
XX
DT  12-OCT-2001  (first entry)
XX
DE  Peptide #5372 encoded by probe for measuring cervical gene expression.
XX
KW  Probe; human; microarray; gene expression; cervical epithelial cell;
KW  cervical cancer.
XX
OS  Homo sapiens.
XX
PN  WO200157278-A2.
XX
PD  09-AUG-2001.
XX
PF  30-JAN-2001; 2001WO-US00670.
XX
PR  04-FEB-2000; 2000US-0180312.
PR  26-MAY-2000; 2000US-0207456.
PR  30-JUN-2000; 2000US-0608408.
PR  03-AUG-2000; 2000US-0632366.
PR  21-SEP-2000; 2000US-0234687.
PR  27-SEP-2000; 2000US-0236359.
PR  04-OCT-2000; 2000GB-0024263.
XX
PA  (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI  Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR  WPI; 2001-488901/53.
XX
PT  Human genome-derived single exon nucleic acid probes useful for
PT  analyzing gene expression in human cervical epithelial cells -
XX
PS  Claim 27; SEQ ID No 23764; 487pp; English.
XX
CC  The present invention relates to human single exon nucleic acid probes
CC  (SENP: see AAI10068-AAI28459). The present sequence is a peptide encoded
CC  by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC  can be used to produce a single exon microarray, which can be used for
CC  measuring human gene expression in a sample derived from human cervical
CC  epithelial cells. By measuring gene expression, the probes are therefore
CC  useful in grading and/or staging of diseases of the cervix, notably
CC  cervical cancer.
CC  Note: The sequence data for this patent did not form part of the printed
CC  specification, but was obtained in electronic format directly from WIPO
CC  at ftp.wipo.int/pub/published_pct_sequences.
```

```
XX
SQ   Sequence   49 AA;


  Query Match             21.5%;  Score 259;  DB 22;  Length 49;
  Best Local Similarity   100.0%;  Pred. No. 5.8e-19;
  Matches   49;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     25 NRPSLKSLKTDNRPEKSKCKPLWGKVFYLDLPSVTISEKLQKDIKDLGG 73
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 nrpslkslktdnrpekskckplwgkvfyldlpsvtiseklqkdikdlgg 49


RESULT   8
AAM31512
ID   AAM31512 standard; Protein; 49 AA.
XX
AC   AAM31512;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Peptide #5549 encoded by probe for measuring placental gene expression.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 27; SEQ ID No 31781; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP:
CC   see AAI31315-AAI57546). The present sequence is a peptide encoded by one
CC   such probe. The probes are useful for producing a microarray for
CC   predicting, measuring and displaying gene expression in samples derived
CC   from human placenta. The probes are useful for antenatal diagnosis of
CC   human genetic disorders.
XX
SQ   Sequence   49 AA;


  Query Match             21.5%;  Score 259;  DB 22;  Length 49;
  Best Local Similarity   100.0%;  Pred. No. 5.8e-19;
  Matches   49;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     25 NRPSLKSLKTDNRPEKSKCKPLWGKVFYLDLPSVTISEKLQKDIKDLGG 73
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 nrpslkslktdnrpekskckplwgkvfyldlpsvtiseklqkdikdlgg 49


RESULT   9
```

```
AAB95297
ID   AAB95297 standard; Protein; 170 AA.
XX
AC   AAB95297;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:17525.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 17525; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   170 AA;


  Query Match             17.2%;  Score 208;  DB 22;  Length 170;
  Best Local Similarity   36.4%;  Pred. No. 4.9e-13;
  Matches   55;  Conservative   22;  Mismatches   44;  Indels   30;  Gaps    4;

Qy     21 KNEKNRPSLKSLKTDNRPEKSKCKPLWGKVFYLDLPSVTISEKLQKDIKDLGGRVEEFLS 80
          | :|| |  :              |  || ||||||:    : |   |: ||| :| |||
Db     35 kcqknspgark------------hpfsgksfyldlpagknlqfltgaiqqlggviegfls 82
```

```
Qy      81 KDISYLISNKKEAKFAQTLGRI-----SPVPSPESAYTAETTSPHPSHDGSSFKSPDTVC 135
           |::||::|:::| | |:: |:      || ||     |:    |  ||   | |  |:|
Db      83 kevsyivssrrevk-aessgkshrgcpspspsevrvetsamvdpkgshprpsrkpvdsvp 141

Qy     136 LSRGKLLVEKAIKDHDFIPSNSILSNALSWG 166
           ||||| |::|||:            | :|||
Db     142 lsrgkellqkair------------nqvswg 160


RESULT  10
AAU03696
ID   AAU03696 standard; Protein; 257 AA.
XX
AC   AAU03696;
XX
DT   12-SEP-2001  (first entry)
XX
DE   Group B Streptococcus antigenic protein, ID-173.
XX
KW   Group B Streptococcus; encapsulated bacterium; therapeutic; sepsis;
KW   meningitis; neonate; antigenic; vaccine; infection; genital tract;
KW   capsid polysaccharide vaccination.
XX
OS   Streptococcus agalactiae.
XX
PN   WO200132882-A2.
XX
PD   10-MAY-2001.
XX
PF   07-SEP-2000; 2000WO-GB03437.
XX
PR   07-SEP-1999;   99GB-0021125.
XX
PA   (MICR-) MICROBIAL TECHNICS LTD.
XX
PI   Le Page RWF,  Wells JM,  Hanniffy SB;
XX
DR   WPI; 2001-316444/33.
DR   N-PSDB; AAS07113.
XX
PT   New polypeptides derived from Streptococcus agalactiae are useful to
PT   provide detection of, and vaccination against, Group B Streptococcus
PT   infections, particularly to prevent infection in neonatals -
XX
PS   Claim 1; Fig 1; 178pp; English.
XX
CC   AAU03601-AAU03722 represent Group B Streptococcus (Streptococcus
CC   agalactiae) amino acid sequences of the invention.  S. agalactiae is an
CC   encapsulated bacterium which is a major pathogen of humans causing sepsis
CC   and meningitis in neonates as well as adults. The S. agalactiae antigenic
CC   polypeptides are used to vaccinate against Group B Streptococcus
CC   infections, particularly to prevent infection in new born children
CC   arising from the maternal genital tract. An immunogenic composition is
CC   useful in the preparation of a medicament for the treatment or
CC   prophylaxis of Group B Streptococcus infection. The invention does not
CC   have the disadvantages of varied response rate associated with prior art
CC   capsid polysaccharide vaccination against Group B Streptococcus.
XX
SQ   Sequence   257 AA;


  Query Match             8.0%;  Score 96;  DB 22;  Length 257;
  Best Local Similarity   23.3%;  Pred. No. 0.18;
  Matches   45;  Conservative   34;  Mismatches   54;  Indels   60;  Gaps    8;

Qy      52 YLDLPSVT-ISEKLQKDIKDLGGRVE---EFLSKDISYLI----------------SNKK 91
           :|  | :  : ||   |:: || :||   | :|  :||::                || |
Db      23 flqtpemakllekrgydvrylgyqvenkleiis--lsyimpvtggfqmkidsgpvhsnsk 80

Qy      92 EAK--------FAQTLGRISPVPSPESAYTAETTSPHPSHDG----------SSFKSPDT 133
             |        :|:: | :  :  |   |   |:|  ||: |          | :
Db      81 ylkqfykalqgyaksngvlelivepfddyqlftssgvpsnqgndnliedftssgyhhdgl 140
```

```
Qy     134 VCLSRGKLLVEKAIKDHDFIPSNSILSN----------ALSWGVKILHIDDIRYYIEQK 182
               || |    :|: : : | ::||:           |:|:|:|:  :         |
Db     141 ttgftgkylswhyvknlegvtsetllssfsktgralvkkamsfgikvrvl---------k 191

Qy     183 KKELYLLKKSSTS 195
           : ||:| |: :||
Db     192 rdelhlfkeitts 204


RESULT  11
AAY49070
ID   AAY49070 standard; Protein; 1435 AA.
XX
AC   AAY49070;
XX
DT   05-JAN-2000  (first entry)
XX
DE   PolC gene product Pol III-L.
XX
KW   Gram positive bacteria; dnaE; dnaX; dnaB; PolC; dnaN; dnaG; helicase;
KW   alpha subunit; DNA polymerase III holoenzyme; gamma subunit; tau subunit;
KW   clamp loader; glue protein; replication; antibiotic.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1033..1047
FT                   /note= "Encoded by GTA"
XX
PN   WO9937661-A1.
XX
PD   29-JUL-1999.
XX
PF   25-JAN-1999;   99WO-US01547.
XX
PR   27-JAN-1998;   98US-0074522.
PR   22-JUL-1998;   98US-0093727.
XX
PA   (UYRQ ) UNIV ROCKEFELLER.
XX
PI   O'Donnell ME,  Zhang D,  Whipple R;
XX
DR   WPI; 1999-590685/50.
DR   N-PSDB; AAZ31004.
XX
PT   New isolated dnaE, dnaX and dnaB genes from Gram positive bacteria,
PT   used to develop screening assays for identifying antibiotic compounds
PT   -
XX
PS   Example 6; Page 25-30; 132pp; English.
XX
CC   This is the PolC gene product Pol III-l of Staphylococcus aureus. The
CC   invention relates to a number of isolated DNA molecules from Gram
CC   positive bacterium, corresponding to dnaE (AAZ31001), dnaX
CC   (AAZ31002),and dnaB (AAZ31003). The PolC, dnaN and dnaG genes
CC   (AAZ31004-Z31006) are also identified. The dnaE gene corresponds to the
CC   alpha subunit of the Escherichia coli, DNA polymerase III holoenzyme,
CC   dnaX corresponds to the gamma and tau subunits, and dnaB corresponds to
CC   the helicase. The alpha subunit is the actual DNA polymerase, the gamma
CC   complex forms the clamp loader and tau is a "glue protein". DnaX encodes
CC   both gamma and Tau, Tau is the product of the full gene, while gamma is
CC   the product of the first two thirds of the gene. The DNA sequences of the
CC   invention can be used to identify agents that inhibit or promote DNA
CC   replication by acting on various parts of the gram positive bacterial DNA
CC   polymerase holoenzyme. The products and methods of the invention can be
CC   used for identifying pharmacological agents or lead compounds for agents
CC   active at the level of a replication protein function, particularly DNA
CC   replication. The agents identified can be used as antibiotics.
XX
SQ   Sequence   1435 AA;


  Query Match             7.8%;  Score 93.5;  DB 20;  Length 1435;
```

```
  Best Local Similarity   25.1%;  Pred. No. 3.8;
  Matches   57;  Conservative   31;  Mismatches   88;  Indels   51;  Gaps   12;

Qy      18 IQVKNEKNRPSLKSLKTDNRP-EKSKCKPLWGKVFYLDLPSVTISEKLQKDIKDLGGRVE 76
           : :|:|   | ::    : |    :  : |:|:    ||| : | ::|:|::| : |
Db     805 vpikdelytprmeganeeirelsyanarklyge----dlpqivi-drlekelksiign-- 857

Qy      77 EFLSKDISYLISN---KKEAKFAQTLGRISPVPSPESAYTAETT--SPHPSH-------- 123
                : ||||    ||       :|    | |   |    | |  :| | |
Db     858 ---gfaviylisqrlvkksldggylvgsrgsvgssfvatmteitevnplpphyicpnckt 914

Qy     124 --------DGSSFKSPDTVCLSRGKLLVEKAIKDHDFIPSNSILSNALSWGVKILHIDDI 175
                   || |  ||  | : |  |    ||:   ||    |      | |:  | |:
Db     915 seffndgsvgsgfdlpdktcetcgapl----ikegqdipfekflgfk---gdkvpdi-dl 966

Qy     176 RYYIEQKKKELYLLKKSSTSVRDGGKRV------GSGAQKTRTGRLK 216
           :  | :         : | |  |  :|      |: |:||  | :|
Db     967 nfsgeyqpn-----ahnytkvlfgedkvfragtigtvaektafgyvk 1008


RESULT  12
AAB31934
ID   AAB31934 standard; Protein; 1435 AA.
XX
AC   AAB31934;
XX
DT   15-MAY-2001  (first entry)
XX
DE   Amino acid sequence of a partial polC polypeptide.
XX
KW   dnaE; Gram positive bacteria; polC; dnaE; holA; holB; dnaX; dnaN; ssb;
KW   dnaG; dnaB; antibiotic; replication; cell growth; cell death;
KW   bacterial infection.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 207
FT                   /note= "encoded by AAC"
FT   Misc-difference 208
FT                   /note= "encoded by GAA"
FT   Misc-difference 952
FT                   /note= "encoded by ACA"
FT   Misc-difference 1035
FT                   /note= "encoded by CGT"
FT   Misc-difference 1147
FT                   /note= "encoded by GAATTC"
FT   Misc-difference 1148
FT                   /note= "encoded by GGT"
FT   Misc-difference 1149
FT                   /note= "encoded by ACA"
FT   Misc-difference 1150
FT                   /note= "encoded by GGA"
FT   Misc-difference 1151
FT                   /note= "encoded by TTC"
FT   Misc-difference 1152
FT                   /note= "encoded by GTG"
FT   Misc-difference 1259
FT                   /note= "encoded by ATG"
FT   Misc-difference 1408
FT                   /note= "encoded by TCT"
XX
PN   WO200109164-A2.
XX
PD   08-FEB-2001.
XX
PF   28-JUL-2000; 2000WO-US20666.
XX
PR   29-JUL-1999;   99US-0146178.
XX
PA   (UYRQ ) UNIV ROCKEFELLER.
XX
PI   O'Donnell ME,  Bruck I,  Zhang D,  Whipple R;
XX
DR   WPI; 2001-147453/15.
DR   N-PSDB; AAF54734.
XX
PT   Isolated DNA molecule from a Gram positive bacterium encoding DNA
PT   replication proteins used to identify compounds which have antibiotic
PT   activity -
XX
PS   Disclosure; Page 35-39; 239pp; English.
XX
CC   The present sequence represents a partial polC polypeptide. The
CC   specification describes DNA molecules from Gram positive bacteria,
CC   which comprise a coding region from a polC, dnaE, holA, holB, dnaX,
CC   dnaN, ssb, dnaG or a dnaB gene. These sequences encode proteins that
CC   replicate the chromosome of Gram positive bacteria. They are used for
CC   sequencing and amplification of DNA and in drug discovery to identify
CC   compounds which have antibiotic activity through interference with
CC   replication. They are used in methods for identifying compounds that
CC   are active at the level of DNA replication and result in arrest of
CC   cell growth or cell death of bacteria to treat bacterial infections in
CC   animals.
XX
SQ   Sequence   1435 AA;


  Query Match             7.8%;  Score 93.5;  DB 22;  Length 1435;
  Best Local Similarity   25.1%;  Pred. No. 3.8;
  Matches   57;  Conservative   31;  Mismatches   88;  Indels   51;  Gaps   12;

Qy      18 IQVKNEKNRPSLKSLKTDNRP-EKSKCKPLWGKVFYLDLPSVTISEKLQKDIKDLGGRVE 76
           : :|:|   | ::    : |    :  : |:|:    ||| : | ::|:|::| : |
Db     805 vpikdelytprmeganeeirelsyanarklyge----dlpqivi-drlekelksiign-- 857

Qy      77 EFLSKDISYLISN---KKEAKFAQTLGRISPVPSPESAYTAETT--SPHPSH-------- 123
                : ||||    ||       :|    | |   |    | |  :| | |
Db     858 ---gfaviylisqrlvkksldggylvgsrgsvgssfvatmteitevnplpphyicpnckt 914

Qy     124 --------DGSSFKSPDTVCLSRGKLLVEKAIKDHDFIPSNSILSNALSWGVKILHIDDI 175
                   || |  ||  | : |  |    ||:   ||    |      | |:  | |:
Db     915 seffndgsvgsgfdlpdktcetcgapl----ikegqdipfekflgfk---gdkvpdi-dl 966

Qy     176 RYYIEQKKKELYLLKKSSTSVRDGGKRV------GSGAQKTRTGRLK 216
           :  | :         : | |  |  :|      |: |:||  | :|
Db     967 nfsgeyqpn-----ahnytkvlfgedkvfragtigtvaektafgyvk 1008


RESULT  13
AAY19935
ID   AAY19935 standard; Protein; 1087 AA.
XX
AC   AAY19935;
XX
DT   19-JUL-1999  (first entry)
XX
DE   B. burgdorferi antigenic protein, t742.aa.
XX
KW   Antigenic protein; vaccine; Lyme disease; infection; detection.
XX
OS   Borrelia burgdorferi.
XX
PN   WO9859071-A1.
XX
PD   30-DEC-1998.
XX
PF   18-JUN-1998;   98WO-US12718.
XX
PR   03-SEP-1997;   97US-0057483.
PR   20-JUN-1997;   97US-0050359.
PR   22-JUL-1997;   97US-0053344.
PR   22-JUL-1997;   97US-0053377.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (MEDI-) MEDIMMUNE INC.
```

```
XX
PI   Choi GH,  Erwin AL,  Hanson MS,  Lathigra R;
XX
DR   WPI; 1999-189980/16.
DR   N-PSDB; AAX61632.
XX
PT   New isolated Borrelia burgdorferi nucleic acids  - used to develop
PT   products for the diagnosis, prevention and treatment of diseases
PT   caused by Borrelia, particularly Lyme disease
XX
PS   Claim 12; Page 125; 275pp; English.
XX
CC   This sequence represents a Borrelia burgdorferi (Bb) protein of the
CC   invention, which is suitable for use in a vaccine. The Bb polypeptides
CC   can be used in vaccines for eliciting protective antibodies to members of
CC   the Borrelia genus, particularly for the use against Lyme disease in
CC   humans and animals. They can be used for preventing or attenuating an
CC   infection caused by a member of the Borrelia genus. The products can also
CC   be used for detection of members of the Borrelia genus.
XX
SQ   Sequence   1087 AA;


  Query Match             7.4%;  Score 89;  DB 20;  Length 1087;
  Best Local Similarity   22.8%;  Pred. No. 7.2;
  Matches   58;  Conservative   45;  Mismatches  101;  Indels   50;  Gaps   15;

Qy      22 NEKNRPSLKSLKTDNRPEKSKCKP---LWGKVF--YLDLPSVTISEKLQKDIKDLG---- 72
           |  |  ||| : :::: |     |   : ||::  |    |  | ::| : : |:
Db     217 nnnnttslkkissnsqkeselsppsqtiigkiyrpy----sylikkelyeilddintgrv 272

Qy      73 ----GRVEEFLSKDIS-------YLISNKKEAKFAQTL-----GRISP--VPSPESAY-- 112
                |::| : | :|        || | |  : :  |       | |  :  |:  |
Db     273 tlgknrlkelikkglsnkfqkvnelienskнkeasnllltlikkdiepnlinipkdpykk 332

Qy     113 -----TAETTSPHPSHD-GSSFKSPDTVCLSRGKLLVEKAIKD-HDFI---PSNSILSNA 162
                 |   |    |   |   |   : |   |   ::|||| ::|:   |:::  |
Db     333 eifqldkedkkpqyledlkskvhsikpidlentksr-qqaikdlneflknnpndaqaskt 391

Qy     163 LSWGVKILHIDDIRYYIEQ-KKKELYLLKKSSTSVRDGG---KRVGSGAQKTRTGRLKKP 218
           |:   || |::|::  :   |  :|   |    :::|     |   : || ::|
Db     392 laqankiqhledlkskvhsikpidlentksrqqaikdlneflknnpndaqasktlaqank 451

Qy     219 FVKVEDMSQSPAVH 232
             :||:     ||
Db     452 iqhledlksk--vh 463


RESULT  14
AAY19934
ID   AAY19934 standard; Protein; 1119 AA.
XX
AC   AAY19934;
XX
DT   19-JUL-1999  (first entry)
XX
DE   B. burgdorferi antigenic protein, f742.aa.
XX
KW   Antigenic protein; vaccine; Lyme disease; infection; detection.
XX
OS   Borrelia burgdorferi.
XX
PN   WO9859071-A1.
XX
PD   30-DEC-1998.
XX
PF   18-JUN-1998;   98WO-US12718.
XX
PR   03-SEP-1997;   97US-0057483.
PR   20-JUN-1997;   97US-0050359.
PR   22-JUL-1997;   97US-0053344.
PR   22-JUL-1997;   97US-0053377.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Choi GH,  Erwin AL,  Hanson MS,  Lathigra R;
XX
DR   WPI; 1999-189980/16.
DR   N-PSDB; AAX61631.
XX
PT   New isolated Borrelia burgdorferi nucleic acids  - used to develop
PT   products for the diagnosis, prevention and treatment of diseases
PT   caused by Borrelia, particularly Lyme disease
XX
PS   Claim 12; Page 124-125; 275pp; English.
XX
CC   This sequence represents a Borrelia burgdorferi (Bb) protein of the
CC   invention, which is suitable for use in a vaccine. The Bb polypeptides
CC   can be used in vaccines for eliciting protective antibodies to members of
CC   the Borrelia genus, particularly for the use against Lyme disease in
CC   humans and animals. They can be used for preventing or attenuating an
CC   infection caused by a member of the Borrelia genus. The products can also
CC   be used for detection of members of the Borrelia genus.
XX
SQ   Sequence   1119 AA;


  Query Match             7.4%;  Score 89;  DB 20;  Length 1119;
  Best Local Similarity   22.8%;  Pred. No. 7.6;
  Matches   58;  Conservative   45;  Mismatches  101;  Indels   50;  Gaps   15;

Qy      22 NEKNRPSLKSLKTDNRPEKSKCKP---LWGKVF--YLDLPSVTISEKLQKDIKDLG---- 72
           |  |  ||| : :::: |     |   : ||::  |    |  | ::| : : |:
Db     249 nnnnttslkkissnsqkeselsppsqtiigkiyrpy----sylikkelyeilddintgrv 304

Qy      73 ----GRVEEFLSKDIS-------YLISNKKEAKFAQTL-----GRISP--VPSPESAY-- 112
                |::| : | :|        || | |  : :  |       | |  :  |:  |
Db     305 tlgknrlkelikkglsnkfqkvnelienskнkeasnllltlikkdiepnlinipkdpykk 364

Qy     113 -----TAETTSPHPSHD-GSSFKSPDTVCLSRGKLLVEKAIKD-HDFI---PSNSILSNA 162
                 |   |    |   |   |   : |   |   ::|||| ::|:   |:::  |
Db     365 eifqldkedkkpqyledlkskvhsikpidlentksr-qqaikdlneflknnpndaqaskt 423

Qy     163 LSWGVKILHIDDIRYYIEQ-KKKELYLLKKSSTSVRDGG---KRVGSGAQKTRTGRLKKP 218
           |:   || |::|::  :   |  :|   |    :::|     |   : || ::|
Db     424 laqankiqhledlkskvhsikpidlentksrqqaikdlneflknnpndaqasktlaqank 483

Qy     219 FVKVEDMSQSPAVH 232
             :||:     ||
Db     484 iqhledlksk--vh 495


RESULT  15
AAR13227
ID   AAR13227 standard; Protein; 700 AA.
XX
AC   AAR13227;
XX
DT   14-OCT-1991  (first entry)
XX
DE   Novel endoglucanase.
XX
KW   Cellulase activity; detergent.
XX
OS   Bacillus spp. NCIMB 40250.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..31
FT                   /label= signal sequence
FT   Protein         32..700
FT                   /label= mature endoglucanase
FT   Cleavage-site   31..32
XX
PN   WO9110732-A.
XX
```

PD 25-JUL-1991.
XX
PF 18-JAN-1991; 91WO-DK00013.
XX
PR 19-JAN-1990; 90DK-0000164.
XX
PA (NOVO) NOVO NORDISK A/S.
XX
PI Jorgensen PL, Schulein M, Hansen C;
XX
DR WPI; 1991-238020/32.
DR N-PSDB; AAQ13001.
XX
PT Enzyme exhibiting cellulase activity from Bacillus sp. - is an
PT endo-glucanase, esp. useful for harshness redn. of cotton-contg.
PT fabrics.
XX
PS Claim 1; Page 80; 96pp; English.
XX
CC The enzyme is produced by a strain of Bacillus spp. NCIMB 40250
CC and exhibits an endoglucanase activity of at least 10 (pref. at
CC least 25) carboxymethyl cellulose (CMC) endoase units per mg total
CC protein under alkaline conditions. It is especially useful as a
CC cellulolytic agent and has been found to be more stable during
CC washing (60 mins. at 40 deg.)in the presence of conventional
CC detergents than a commercial cellulase preparation. It may also
CC show increased storage stability in liq. detergents contg.
CC proteases. The sequence was deduced from the DNA (AAQ13001).
CC See also AAR13228 and AAR13229.
XX
SQ Sequence 700 AA;

Query Match 7.3%; Score 88; DB 12; Length 700;
Best Local Similarity 23.9%; Pred. No. 4.8;
Matches 54; Conservative 35; Mismatches 83; Indels 54; Gaps 12;

```
Qy      20 VKNEKNRPSLKSLKTDNRPEKSKCKPLWGKV-FYLDLPSVTISEKLQKDIKDLGGRVEEF 78
           |  :|| |  :|        ::   | | || |  : |       || |: |  |:|  :
Db     142 vsrdkngpvdes--------etapsprwdkvefaknapf-----slqpdlnd--gqv--y 184

Qy      79 LSKDISYLISNKKEAKFAQTLGRISPVPSPESAYTAETTSPHPSHDGSSFKSPDTVCLSR 138
           : :::::|::    |  :  :   |     | |  :||   ||       :: : |  |
Db     185 mdeevnflvnrygnaststgikays--ldnepalwset---hprihpeqlqaaelvaks- 238

Qy     139 GKLLVEKAIKDHDFIPSNSILSNALSWGVKILHIDD-------------IRYYIEQKKK 184
             : : ||:|: |  |   |   ||      | : |              | ||::| |
Db     239 --idlskavknvd--phaeifgpalygfgaylslqdapdwpslqgnyswfidyyldqmkn 294

Qy     185 ELYLLKKSSTSVRD---------GGKRV---GSGAQKTRTGRLKKP 218
             |    | |         ||:|:   |:|   |:  |:: |
Db     295 ahtqngkrlldvldvhwypeaqgggqrivfggagnidtqkarvqap 340
```

Search completed: December 27, 2001, 16:53:20
Job time: 520 sec